DTU Library

Effects of slow- and fast-acting compression on hearing impaired listeners' consonant-vowel identification in interrupted noise

Kowalewski, Borys; Zaar, Johannes; Fereczkowski, Michal; MacDonald, Ewen; Strelcyk, Olaf ; May, Tobias; Dau, Torsten

Published in:
Proceedings of the International Symposium on Auditory and Audiological Research: Adaptive Processes in Hearing

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Kowalewski, B., Zaar, J., Fereczkowski, M., MacDonald, E., Strelcyk, O., May, T., & Dau, T. (2017). Effects of slow- and fast-acting compression on hearing impaired listeners' consonant-vowel identification in interrupted noise. In S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, & T. Poulsen (Eds.), *Proceedings of the International Symposium on Auditory and Audiological Research: Adaptive Processes in Hearing* (Vol. 6). The Danavox Jubilee Foundation. Trends in Hearing Vol. 22

Effects of slow- and fast-acting compression on hearing-impaired listeners' consonant-vowel identification in interrupted noise

BORYS KOWALEWSKI,[1,*] JOHANNES ZAAR[1], MICHAL FERECZKOWSKI[1], EWEN N. MACDONALD[1], OLAF STRELCYK[2], TOBIAS MAY[1], AND TORSTEN DAU[1]

[1] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[2] *Sonova U.S. Corporate Services, Warrenville, IL, USA*

There is conflicting evidence about the relative benefit of slow- and fast-acting compression for speech intelligibility. It has been hypothesized that fast-acting compression improves audibility at low signal-to-noise ratios (SNRs) but may distort the speech envelope at higher SNRs. The present study investigated the effects of compression with nearly instantaneous attack time but either fast (10 ms) or slow (500 ms) release times on consonant identification in hearing-impaired listeners. Consonant-vowel speech tokens were presented at several presentation levels in two conditions: in the presence of interrupted noise and in quiet (with the compressor "shadow-controlled" by the corresponding mixture of speech and noise). These conditions were chosen to disentangle the effects of consonant audibility and noise-induced forward masking on speech intelligibility. A small but systematic intelligibility benefit of fast-acting compression was found in both the quiet and the noisy conditions for the lower speech levels. No negative effects of fast-acting compression were observed when the speech level exceeded the level of the noise. These findings suggest that fast-acting compression provides an audibility benefit in fluctuating interferers as compared to slow-acting compression, while not substantially affecting the perception of consonants at higher SNRs.

INTRODUCTION

It is widely accepted that due to the limited dynamic range of levels perceived by hearing-impaired (HI) listeners, some sort of level-dependent amplification is required to compensate for hearing loss. The majority of modern hearing aids apply dynamic-range compression (see Souza, 2002, and Edwards, 2004, for reviews). In such systems, the gain is determined based on one or more level-estimation circuit(s), characterized by attack and release time constants. The most commonly used attack times have values below 10 ms (Jenstad and Souza, 2005) in order to quickly reduce the gain in response to loud sounds. However, the optimal speed of gain recovery, i.e.,

*Corresponding author: bokowal@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.
 ISBN: 978-87-990013-6-1.

the release time, is still a subject of discussion. Shorter release times allow more gain to be applied to the low-intensity speech components (e.g., consonants) that follow other, high-intensity components (e.g., vowels) or noise bursts. This increased gain can potentially improve audibility and reduce the amount of forward masking, which in turn might lead to an improved speech recognition performance in HI listeners (Souza and Bishop, 1999; Edwards, 2002; Desloge *et al.*, 2010; Jenstad and Souza, 2005). On the other hand, with a very short release time, the gain follows the fast fluctuations of the signal, effectively reducing the temporal contrast. The temporal characteristics of the speech signal provide important cues for speech intelligibility, especially for HI listeners (Souza *et al.*, 2015). Temporal envelope distortion introduced by fast-acting amplification might therefore lead to a decrement in recognition performance. It is possible that optimal performance would be achieved if the time constants were adapted dynamically according to the current signal-to-noise ratio (SNR). For example, May *et al.* (2017) proposed a blind broadband-SNR estimator (based on the speech and noise power spectrum density), which could be applied in hearing aids. However, the relation between the optimal release time and SNR in connection to speech intelligibility is not yet known.

In the present study, it is hypothesized that potential negative effects of short release times will be more pronounced at higher SNRs, where audibility and masking are less of a concern and the compression is driven mostly by the speech signal. On the other hand, the additional gain applied by the fast-acting system is expected to provide an increasing benefit as the SNR decreases. To test these ideas, stimuli were designed to maximize the effects of compression release time. The noise consisted of high-intensity bursts, separated by silent gaps and had very sharp onsets and offsets. Short consonant-vowel (CV) tokens were used and listeners were asked to report the initial consonant – a speech component that typically has a low intensity. The temporal onset of the CV token relative to the noise was controlled and chosen based on a previous study (Zaar *et al.*, 2017). A wide range of SNRs and compression release times were tested in order to capture the potential interaction between the two factors.

METHODS

Listeners

Twelve young, normal-hearing (NH) listeners aged between 19 and 26 years (average age: 21.7 years) completed the task in the unaided condition. They all had pure-tone thresholds lower than 20 dB HL in the 250 to 8000 Hz range. The aided conditions were completed by nine older HI listeners aged between 66 and 77 years (average age: 71.4 years). Their hearing losses ranged from mild to moderately-severe losses and were most prominent at the high frequencies.

Stimuli

The target speech consisted of 15 consonant-vowel (CV) tokens: */bi, di, fi, gi, hi, ji, ki, li, mi, ni, pi, si, ʃi, ti, vi/* spoken by one male and one female talker (30 utterances in total), used previously by Zaar and Dau (2015). Four presentation levels were used:

45, 55, 65, and 75 dB SPL. In the aided conditions, these were the levels at the *input* to the amplification system. In each condition, each utterance was presented five times to the listeners.

The noise was composed of five 100-ms long bursts, separated by 100-ms silent gaps (corresponding to a 5-Hz repetition rate). White noise was chosen as a carrier in order to maximize masking of high-frequency consonants. The sound pressure level was 65 dB, defined as the level of the noise bursts at the input to the hearing aid simulator. The onset of the CV token was positioned 25 ms into the silent gap after the third noise burst, as shown schematically in Fig. 1. The instantaneous SNR was therefore infinite. The broadband SNR values are still reported for consistency with previous literature. They are defined as the difference between the sound pressure level of the token and the preceding noise burst.

Thirty noise waveforms (one per utterance) were pre-generated and stored as wav-files. Each utterance was always presented with the same noise recording. This was done in order to limit the across-repetition variability due to the random fluctuations in the Gaussian noise carrier, whilst preventing noise-learning effects that could occur if only one noise-waveform was used for all utterances (see Zaar and Dau, 2015).

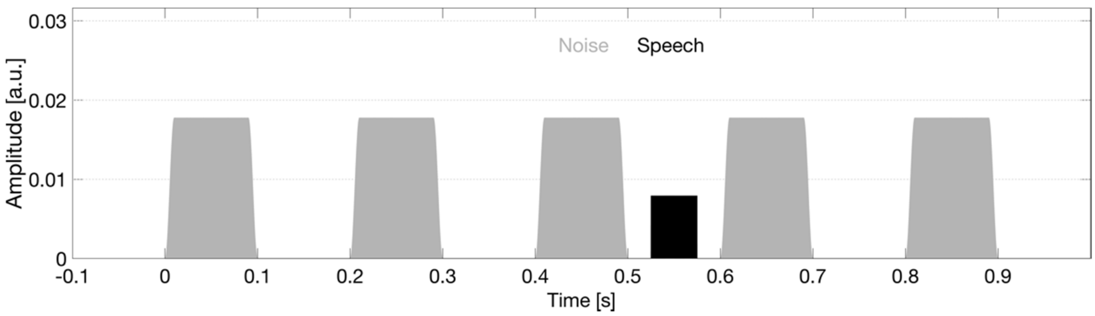

Fig. 1: A schematic representation of the stimulus time-course.

Amplification

For the HI listeners, the stimuli were pre-processed using a hearing-aid simulator with eight independent compression channels, implemented in MATLAB. The insertion gain was applied to the signals presented monaurally over Sennheiser HD650 headphones. The gain was frequency-dependent, based on the NAL-NL2 target (using the *Slow* setting, which yields more aggressive amplification, cf. Keidser *et al.*, 2011) for the N2 audiogram (Bisgaard *et al.*, 2010). This audiogram was chosen because it was most representative of the participants' hearing losses. Thus, compression ratios were the same for all listeners. However, in order to maximize audibility for each participant, the linear part of the gain (gain applied to stimuli below the compression threshold) was determined based on the individual audiogram.

The compression thresholds (kneepoints) were also frequency-dependent and calibrated such that each channel went into compression when the level of a broadband (white-noise) input exceeded 50 dB SPL. The attack time (of the level-detector circuit, or the so-called *RC time constants*, Kates, 1993) was always 5 ms. The release time depended on the amplification condition. It was 10 ms in the *fast* compression condition and 500 ms in the *slow* condition. The third condition was *linear*, which used the same maximum gain values but a compression ratio of 1:1 (i.e., no compression) and null attack and release times. It thus simulated an "idealized" hearing aid that never applies compression and provides the maximum possible amplification. Such high gain is unrealistic for high-intensity inputs, as it would be excessively loud. Thus, this condition served as a baseline for the behavior of compression systems, but only for lower-intensity speech inputs – 45 and 65 dB SPL in quiet (see "Experimental conditions" below).

In all conditions, the level-detection circuit of the compression and the resulting gain were driven by the mixture of speech and noise. Thus, the gain applied to the clean speech in the quiet condition was not controlled by the clean speech signal but rather *shadow-controlled* by the mixture. This setup allowed the investigation of the effects of the gain fluctuations (resulting from the presence of the interrupted noise) on the CV token without actually presenting the interferer to the listeners' ears.

Experimental conditions

The NH listeners were tested unaided while HI listeners were always presented with amplified stimuli. Slow and fast compression were tested in all conditions, while linear amplification was tested only in a limited number of conditions. In quiet, the compressed stimuli were always shadow-filtered with the corresponding mixture of speech and noise. An overview of all experimental conditions is shown in Table 1.

	Speech input level (dB SPL)	NH	HI		
		Unaided	Linear	Slow	Fast
Quiet (shadow-filtered)	45	x	x	x	x
	65	x	x	x	x
	75	-	-	x	x
Noise	45	x	-	x	x
	55	x	-	x	x
	65	x	-	x	x
	75	x	-	x	x

Table 1: Summary of experimental conditions: speech-noise configurations and amplification used.

RESULTS

For the quiet and noisy data sets, separate linear mixed-effects models were used with the fixed factors *speech level* and *amplification type* and random factor *listener.* Backwards elimination of non-significant effects was performed (Kuznetsova *et al.*, 2015) and the final model was used to establish significance between the results obtained with each *amplification type* at each *speech level*.

The distribution of the model residuals for the data in quiet deviated from normal (it was "light-tailed"). Therefore, these data were RAU-transformed before further analysis. The transformation was not necessary for the data in noise (the distribution of residuals was much closer to normal), so only the non-transformed data are reported for consistency.

Consonant recognition in quiet

The average consonant recognition rates for the stimuli presented in quiet are shown in Fig. 2. It can be observed that the unaided NH listeners achieved recognition rates close to 100% at both speech input levels, whereas the aided HI listeners performed much worse in all conditions and achieved maximum recognition rates of about 87% at 75 dB SPL. Significant differences were found between all amplification types for the lowest speech input level (45 dB SPL). The best recognition rate of 55% was achieved with linear amplification (that provided maximum possible gain), followed by fast (46%) and slow compression (34%).

For the 65 dB SPL speech input, no significant differences between amplification types were observed. Between 65 and 75 dB SPL, there was a slight increase in performance with slow compression but no significant change with fast compression (possibly due to ceiling effects). At 75 dB SPL, slow compression yielded, on average, slightly higher recognition rates than fast compression, but the difference was not significant.

Consonant recognition in noise

The recognition rates in noise are shown in Fig. 3. NH listeners achieved recognition rates of 95% for speech levels of 65 and 75 dB SPL (corresponding to SNRs of 0 and +10 dB). The rate decreased to 73% at 45 dB SPL (SNR −20 dB). Aided HI listeners achieved the highest recognition rate of 77% at 75 dB SPL. For speech input levels of 45, 55, and 65 dB SPL, the recognition rates observed with fast compression were 7-9% higher than with slow compression, with all differences being statistically significant ($p < 0.001$). At 75 dB SPL, slow compression yielded slightly higher recognition performance than fast, but the difference was not statistically significant.

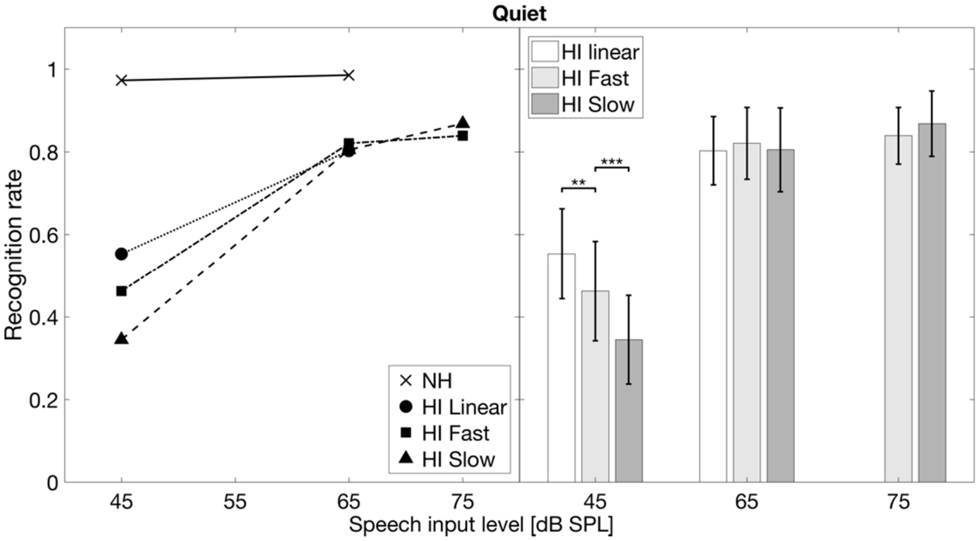

Fig. 2: Averaged consonant recognition rates for speech tokens in quiet, “shadow-controlled” by the mixture of speech and noise. Left: normal-hearing (NH) unaided and hearing-impaired (HI) aided with three types of amplification. Right: Only the HI data replotted. The error bars indicate +/- one standard deviation. The significance levels are: ** 0.01, *** 0.001.

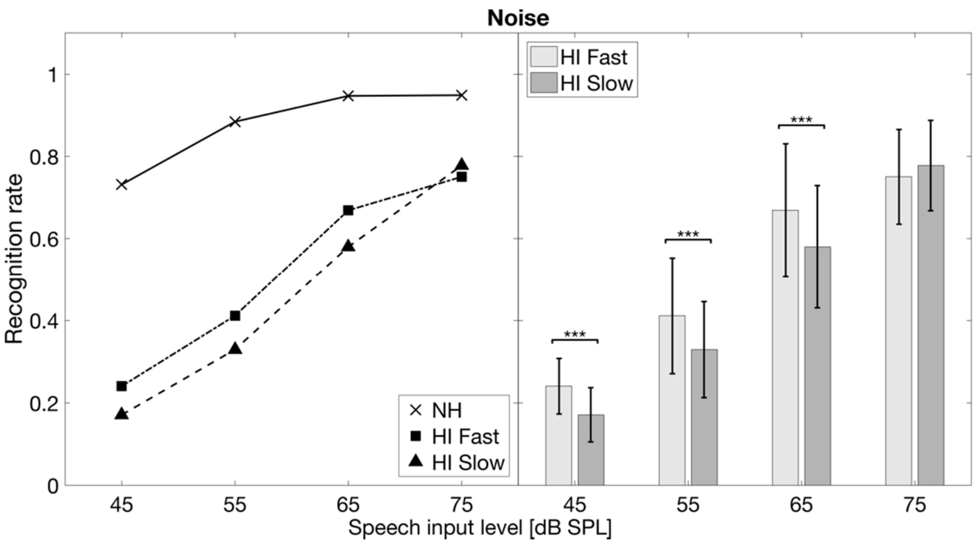

Fig. 3 The same as Fig 2. but presented in 5-Hz interrupted Gaussian noise (noise level: 65 dB SPL).

DISCUSSION

In the quiet condition, the consonant recognition rates at low speech input levels strongly depended on the amplification type. The best performance was obtained with linear amplification and fast compression, which provide higher gain and thus better audibility than slow compression. On the other hand, slow compression induced a small increase in performance between 65 and 75 dB SPL, while in the case of fast compression a ceiling effect was observed. Moreover, slow compression seemed to lead to a better average performance at 75 dB SPL, but the difference was small and not statistically significant.

In noise, fast-acting compression led to higher recognition rates for speech levels of up to 65 dB SPL, corresponding to a broadband SNR of 0 dB. Similar to the results in quiet, the performance at 75 dB SPL (+10 dB SNR) tended to be better with slow compression, but the effect was small and not significant. Overall, there is thus no statistically significant evidence for negative effects of fast compression (e.g., due to temporal envelope distortion) on consonant recognition performance at higher speech input levels. However, possible ceiling effects in the HI listeners' data may be a confounding factor here.

Effects on the recovery from forward masking

In quiet, the relative benefit of fast vs. slow compression decreased from 12% at 45 dB SPL to 2% at 65 dB SPL (SNRs −20 and 0 dB). In noise, the benefit increased from 7 to 9% between speech input levels of 45 and 65 dB SPL. As the compressor was shadow-controlled by the mixture of speech and noise when being applied to speech in quiet, it behaved identically in both conditions such that the only difference between the quiet and noise conditions was the presence of the noise. Therefore, an explanation for the above observation may be that the higher gain provided to the speech token by fast-acting compression improved the recovery from the noise-induced forward masking, at least at SNRs close to 0 dB (i.e., speech levels close to 65 dB SPL).

CONCLUSIONS AND OUTLOOK

A small but systematic benefit of fast-acting compression was found both in the quiet and the noisy conditions for speech levels below 65 dB SPL (0 dB SNR in noise). Despite potentially detrimental speech envelope distortions, no significant negative effects of fast-acting compression were observed when the speech level exceeded the level of the noise. These findings suggest that fast-acting compression provides an audibility benefit and, possibly, an improved recovery from forward masking in fluctuating interferers as compared to slow-acting compression, while not substantially compromising the perception of short CV tokens at higher SNRs.

It is yet to be investigated whether these effects persist in more realistic conditions, i.e., with longer speech stimuli (multi-syllable words, sentences) in fluctuating interferers with softer onsets/offsets.

The findings from this study and prospective future studies may help design SNR-dependent amplification strategies and individualized hearing-aid fitting strategies.

The potential use of blind SNR estimation for hearing-aid applications has been investigated in May *et al.* (2017). The output of such an estimator could be used to dynamically manipulate compression parameters in real-time and will be the subject of future investigations.

REFERENCES

Bisgaard, N., Vlaming, M., and Dahlquis, M. (**2010**). "Standard audiograms for the IEC 60118-15 measurement procedure," Trends Amplif., **14**, 113-120, doi: 10.1177/1084713810379609

Desloge, J.G., Reed, C.M., Braida, L.D., Perez, Z.D., and Delhorne, L.A. (**2010**). "Speech reception by listeners with real and simulated hearing impairment: effects of continuous and interrupted noise," J. Acoust. Soc. Am., **128**, 342-359, doi: 10.1121/1.3436522.

Edwards, B. (**2002**). "Signal processing, hearing aid design and the psychoacoustic Turing test," IEEE ICASSP. doi: 10.1109/ICASSP.2002.5745533

Edwards, B. (**2004**). "Hearing aids and hearing Impairment," in *Speech Processing in the Auditory System*, Edited by R.R. Fay and S. Greenberg (Springer, New York), pp. 339-421. doi: 10.1007/b97399

Jenstad, L., and Souza, P. (**2005**). "Quantifying the effect of compression hearing aid release time on speech acoustics and intelligibility," J. Speech Lang. Hear. Res., **48**, 651-667, doi: 10.1044/1092-4388(2005/045)

Kates, J. (**1993**). "Optimal estimation of hearing-aid compression parameters," J. Acoust. Soc. Am., **94**, 1-12.

Keidser, G., Dillon, H., Flax,. M, Ching, T., and Brewer, S. (**2011**). "The NAL-NL2 prescription procedure," Aud. Res., **1**, 88-90. doi: 10.4081/audiores.2011.e24

Kuznetsova, A., Brockhoff P.B., and Bojesen Christensen, R.H. (**2015**). "Package 'lmerTest'", *R package version* 2.0.

May, T., Kowalewski, B., Fereczkowski, M., and MacDonald E.N. (**2017**). "Assessment of broadband SNR estimation for hearing-aid applications," Proceedings of IEEE ICASSP, 231-235. doi: 10.1109/ICASSP.2017.7952152

Souza, P., and Bishop, R.D. (**1999**). "Improving speech audibility with wide dynamic range compression in listeners with severe sensorineural loss," Ear Hearing, **20**, 461-470. doi: 10.1097/AUD.0b013e3181aec5bc

Souza, P. (**2002**). "Effects of compression on speech acoustics, intelligibility and sound quality," Trends Amplif., **6**, 131-165. doi: 10.1177/108471380200600402.

Souza, P.E., Wright, R.A., Blackburn, M.C., Tatman, R., and Gallun, F.J. (**2015**). "Individual sensitivity to spectral and temporal cues in listeners with hearing impairment," J. Speech Lang. Hear. Res., **58**, 520-534. doi: 10.1044/2015_JSLHR-H-14-0138.

Zaar, J., and Dau, T. (**2015**). "Sources of variability in consonant perception of normal-hearing listeners", J. Acoust. Soc. Am., **138**, 1253-1267. doi: 10.1121/1.4928142.

Zaar, J., Kowalewski, B., and Dau, T. (**2017**) "Effects of non-stationary noise on consonant identification," Poster presented at the International Symposium on Auditory and Audiological Research, Nyborg, Denmark.